

**STATISTICAL
CONCEPTS
FOR
ATTORNEYS**

A REFERENCE GUIDE

Wayne C. Curtis

STATISTICAL CONCEPTS FOR ATTORNEYS

A REFERENCE GUIDE

Wayne C. Curtis



QUORUM BOOKS
WESTPORT, CONNECTICUT
LONDON, ENGLAND

Library of Congress Cataloging in Publication Data

Curtis, Wayne C.

Statistical concepts for attorneys.

Bibliography: p.

Includes index.

1. Law—United States—Statistical methods.

2. Statistics. I. Title.

KF320.S73C87 1983 349.73'072 82-24068

ISBN 0-89930-033-2 (lib. bdg.) 347.30072

Copyright © 1983 by Wayne C. Curtis

All rights reserved. No portion of this book may be reproduced, by any process or technique, without the express written consent of the publisher.

Library of Congress Catalog Card Number: 82-24068

ISBN: 0-89930-033-2

First published in 1983 by Quorum Books

Greenwood Press

A division of Congressional Information Service, Inc.

88 Post Road West, Westport, Connecticut 06881

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Copyright Acknowledgment

Grateful acknowledgment is given for the permission to use excerpts from "Do Polls Tell the Story," a copyrighted interview in *U.S. News & World Report*, October 5, 1964. Used with the permission of *U.S. News & World Report*.

PREFACE

Societal changes during the twentieth century have been marked by an increasing utilization of statistics in nearly every field of endeavor. During the past two decades, statistics has become even more important in conceptualizing new relationships and in evaluating patterns of difference. Today such diverse fields as agriculture, natural science, social science, engineering, nursing, and psychology rely on statistics to express relationships and to evaluate products.

A growing proportion of legal issues involves the use of numerical data. Statistics, which deals with the analysis and interpretation of numerical data, is a useful tool to aid the courts in decision making. This book is a basic reference for attorneys who have limited or no prior knowledge of the field. In addition to legal practitioners in need of a basic reference, the user may also be a law school professor who wishes to review a specific concept or a student in the classroom who utilizes the book as a supplemental text. Nevertheless, the primary purpose is to introduce elementary statistics concepts to the practicing attorney; the focus of the material is on a broad range of legal applications. Attention is devoted to explaining how attorneys may encounter various statistical concepts—and how they may understand and effectively use them—in everyday situations. There is no attempt to provide an exhaustive discussion of each topic; rather the thrust is toward introducing and illustrating basic concepts. For those who wish to pursue additional reading on various topics, selected references are provided at the end of each chapter.

The material assumes no more than a working knowledge of basic algebra. Concepts and principles are developed without tedious and sophisticated derivations. Throughout the book the emphasis is on introducing topics on an intuitive level and then developing these topics into formal statistical concepts. Every attempt has been made to avoid purely academic presentations of the type found in college textbooks.

I believe attorneys will see inference as the objective of statistics. The average practicing attorney understands why decision procedures based on statistical inference function better than those based on intuition. Thus, an awareness of the role probability plays in drawing inferences is essential. Consequently, I have attempted to fuse the theories of probability and inference throughout the book.

In addition to the selected references noted earlier, each chapter contains actual cases and problems illustrating the specific legal application of the concept under discussion. To the extent possible, full citations of cases where certain statistical techniques have been applied are presented.

I am indebted to many persons for their interest and encouragement in the preparation of this manuscript. In particular, appreciation is expressed to my colleague, William H. Smith, statistician nonpareil, for his careful review and constructive criticism of the work; to Lowell E. Wilson, my coauthor in several publishing endeavors, for his invaluable assistance in providing reference materials for me; to my former colleague, Donald C. Hines, for graciously allowing me to draw upon some of his unpublished materials for Chapter 13; and especially to my wife, Seretha, for typing drafts of the manuscript as well as encouraging me to continue with the project. Of course, I assume full responsibility for any errors, omissions, or inconsistencies.

CONTENTS

<i>Figures</i>	<i>xi</i>
<i>Tables</i>	<i>xiii</i>
<i>Preface</i>	<i>xvii</i>

PART I—STATISTICAL DESCRIPTION

Chapter 1: Uses of Statistics in Litigation	3
<i>Statistical Awareness</i>	3
<i>The Role of Statistics in Litigation</i>	4
<i>A Note to the Reader</i>	6
Chapter 2: Data Organization	9
<i>Frequency Distributions</i>	9
<i>Histograms</i>	12
<i>Other Graphic Methods</i>	12
<i>General Purpose Tables</i>	15
<i>Illustrative Uses of Frequency Distributions and Histograms</i>	16
<i>Selected References</i>	20
Chapter 3: Measures of Central Tendency	21
<i>The Mode</i>	21
<i>The Median</i>	22
<i>The Arithmetic Mean</i>	25
<i>Summary of the Mean, the Median, and the Mode</i>	29
<i>Illustrative Uses of the Arithmetic Mean</i>	29
<i>Selected References</i>	32
Chapter 4: Measures of Dispersion	33
<i>Range</i>	33
<i>Variance and Standard Deviation</i>	34

<i>Illustrative Uses of Sampling</i>	105
<i>Selected References</i>	106
Chapter 10: Estimation	107
<i>Point and Interval Estimates</i>	107
<i>Confidence Intervals for Means</i>	108
<i>Confidence Intervals for Proportions</i>	110
<i>Sample Size</i>	110
<i>Illustrative Uses of Estimation</i>	113
<i>Selected References</i>	115
Chapter 11: Inference: Tests of Significance	117
<i>Rationale for Hypothesis Testing</i>	117
<i>Procedure for Hypothesis Testing</i>	118
<i>Errors in Hypothesis Testing</i>	122
<i>Testing Differences between Means</i>	123
<i>Testing Differences between Proportions</i>	133
<i>Illustrative Uses of Statistical Inference</i>	137
<i>Selected References</i>	140
Chapter 12: The Chi-Square Distribution	141
<i>The Chi-Square Distribution</i>	141
<i>Chi-Square Test—Sampling Frequencies against</i> <i>Presupposed Frequencies</i>	142
<i>Contingency Tables</i>	144
<i>Illustrative Uses of Chi-Square</i>	148
<i>Selected References</i>	151
Chapter 13: Correlation and Regression	153
<i>Correlation</i>	154
<i>Simple Linear Regression Analysis</i>	159
<i>Multiple Regression</i>	165
<i>Illustrative Uses of Correlation and Regression</i>	167
<i>Selected References</i>	170
Chapter 14: Index Numbers	171
<i>Kinds of Index Numbers</i>	172
<i>Construction of Index Numbers of Prices</i>	172
<i>Some Common Index Numbers</i>	176
<i>Problems of Index Number Construction</i>	181
<i>Illustrative Uses of Index Numbers</i>	182
<i>Selected References</i>	186
Appendix: Statistical Tables	187
Index	227

FIGURES

2.1 Histogram of Ages of 50 Employees	13
2.2 Frequency Polygon of Ages of 50 Employees	13
2.3 Bar Chart Showing Distribution of 50 Employees by Sex	14
2.4 Annual Percentage Increases in Rental Charges, by Type of Property, Fairhope Single Tax Corporation	17
2.5 Histogram of Annual Percentage Increases in Rental Charges to Rural Lessees, Fairhope Single Tax Corporation	17
2.6 Annual Percentage Increase in Rentals to Five Lessees, Fairhope Single Tax Corporation	18
3.1 Histogram of Ages of 50 Employees Showing Median	24
4.1 Symmetrical and Skewed Distributions	39
5.1 Venn Diagram, Education Level and Sex	47
6.1 Discrete Probability Distribution of the Experiment Involving Five Tosses of a Fair Coin	61
7.1 The Normal Distribution	72
7.2 Areas under the Normal Curve	73
7.3 Probability of x Being Greater than 60	75
7.4 Probability of x Being Less than 40	76
7.5 Illustrations Using the Standardized Normal Distribution	76
7.6 Binomial Distribution Approximated by a Normal Curve	78
9.1 Use of the Normal Distribution to Illustrate Feed Weight Problem	102
11.1 Two-Tailed Test, Areas of Acceptance and Rejection, $\alpha = .01$	121
11.2 Rejection Area for One-Tailed Test, $\alpha = .05$	122
11.3 Test of Difference between Population Mean and Sample Mean	125
11.4 Test of Difference between Population Mean and Sample Mean	126
11.5 Test of Difference between Mean Earnings of Male and Female Employees	130

12.1 Typical Chi-Square Distribution	142
13.1 Scatter Diagram of Test Scores versus Sales Performance	156
13.2 Regression Line of y on x	159

TABLES

2.1	Tabulation of Ages of 50 Employees, January 1983	10
2.2	Tabulation of Ages of 50 Employees, Arranged in Order of Magnitude, January 1983	10
2.3	Frequency Distribution of Ages of 50 Employees, January 1983	11
2.4	Cumulative Frequency Distribution of Ages of 50 Employees, January 1983	11
2.5	Distribution of 50 Employees by Sex, January 1983	12
2.6	Per Capita Tax Revenues, by Type of Tax, Selected Southern States, 1980	15
2.7	Segregation by Types of Jobs	19
3.1	Ages of 50 Employees, Arrayed in Order of Magnitude, January 1983	22
3.2	Frequency Distribution of Ages of 50 Employees, January 1983	22
3.3	Computation of Weighted Arithmetic Mean of Gasoline Prices at Ten Locations, January 1983	28
3.4	Frequency Distribution of Ages of 50 Employees, January 1983	29
3.5	Racial Distribution and Salary of Employees by Department, as of August 12, 1971	30
4.1	Billable Hours and Mode, Median, and Mean Hours for Two Cases, by Attorney, January 1983	34
4.2	Computation of Variance and Standard Deviation, Ungrouped Data, Six Attorneys Working on Case A, January 1983	36
4.3	Computation of Variance and Standard Deviation, Grouped Data, Average Age of Employees, January 1983	37
5.1	Frequency Table, by Level of Education and Sex, 100 Veniremen, January 1983	47
5.2	Probability Table, by Level of Education and Sex, 100 Veniremen, January 1983	50

6.1 Probabilities for the Possible Outcomes of an Experiment Consisting of Five Tosses of a Fair Coin	61
6.2 Computation of the Expected Value, Number of Heads, Five Tosses of a Fair Coin	62
6.3 Computation of the Variance, Number of Heads, Five Tosses of a Fair Coin	63
6.4 Binomial Probability Equality and Inequality Symbols and Relationships	66
6.5 Binomial Probability Distribution for $n = 5$, $p = .50$	66
8.1 General Format of Decision Theory Model	83
8.2 Payoff Table, Law Firm Example, January 1983	83
8.3 Payoff Table with Expected Monetary Value, Law Firm Example, January 1983	84
8.4 Opportunity Loss Table with Expected Opportunity Loss, Law Firm Example, January 1983	85
8.5 Estimated Present Value of Lost Earnings among Various Occupational Groups, Daniel Reuben Brantley	87
9.1 Table of Random Numbers	93
9.2 Average Hourly Rates as Reported by 75 Law Firms, January 1983	98
11.1 Errors in Hypothesis Testing	123
11.2 Results of Examinations for Assistant Principal, Day Elementary and Junior High Schools, by Race, New York	138
12.1 Observed and Expected Levels of Income, Listenership Survey, January 1983	143
12.2 Cross-Tabulation of Responses, by Sex of Respondent, January 1983	145
12.3 Test of Independence, Sex versus Name Association, January 1983	146
12.4 Cross-Tabulation of Responses, by Educational Level of Respondent, January 1983	147
12.5 Test of Independence, Educational Level versus Name Association, January 1983	147
12.6 Observed and Expected Showings, by Day of the Week and Number of Showings, Pike Theater, July 1977 to March 1978	149
12.7 Total and Expected Showings, by Month and Number of Showings, Pike Theater, January 1977 to March 1978	150
13.1 Employee Test Scores and Employee Sales Performance, 20 Employees, January 1983	155
13.2 Correlation between Employee Test Scores and Employee Sales Performance, 20 Employees, January 1983	157

13.3 Analysis of Variance for the Test Score–Sales Performance Regression Model	164
13.4 Beef Consumption and Prices, Pork Prices, Personal Income, and Population, United States, 1964–80	166
13.5 Analysis of Variance for the Beef Demand Regression Model	167
14.1 Laspeyres' Price Indexes for Five Cost Items, 1975–82 (1975 = 100)	175
14.2 Consumer Price Index for Selected Items, All Urban Consumers, 1981–82 (1967 = 100)	178
14.3 Producer Price Indexes for Selected Items, by Stage of Processing, 1981–82 (1967 = 100)	179
14.4 Indexes of Industrial Production for Selected Items, 1981–82 (1967 = 100)	180
A.1 Cumulative Binomial Probability Distribution	189
A.2 Areas for the Standard Normal Probability Distribution	219
A.3 Critical Values of the t Distribution, One-Tail Area	220
A.4 Critical Values of the Chi-Square Distribution	221
A.5 Critical Values of the F Distribution	222

PART I

Statistical Description

CHAPTER 1

USES OF STATISTICS IN LITIGATION

Suppose one takes a well-shuffled pack of cards, it is quite likely or not unlikely that the top card will prove to be a diamond: the odds are only three to one against; but most people would not say that it is quite likely to be the nine of diamonds for the odds are then fifty-one to one against. On the other hand I think that most people would say that there is a serious possibility or a real danger of its being turned up first and, of course, it is on the cards.

—Lord Reid¹

Contrary to Lord Reid's statement regarding probabilities, statistics is perhaps one of the most misunderstood and misused tools available to attorneys. In many instances, statistics and its applications are popularly misunderstood to include only collecting and analyzing numbers. Nothing could be further from reality. When statistical analyses are performed by those trained in the discipline, the methods are essentially those common to all fields of scientific empiricism.

STATISTICAL AWARENESS

The confusion among attorneys results primarily from the fact that *statistics* has two separate and distinct meanings. *Descriptive statistics* refers to collecting, organizing, and analyzing data. *Inferential statistics* is the drawing of conclusions from data that have been collected, organized, and analyzed. To illustrate this distinction, the arithmetic mean of a sample of observation is descriptive; if it is used to estimate the arithmetic mean of a larger group from which the sample was drawn, inferential techniques are involved.

Most attorneys will see inference as the objective of statistics. The foundation of statistical inference is probability theory. The object of statistical inference is to use information obtained from samples collected through the process of observation to draw conclusions about populations. Using probability theory to solve societal problems on a large-scale basis, however, is of relatively recent origin but is of increasing importance.

Lawyers, judges, and legal scholars can no longer ignore the field of statistics. One of the primary factors in this is the increasing intrusion of the federal government into the economy through various forms of regulation. Within recent years, courts have confronted larger and larger numbers of technical disputes involving problems such as air and water quality, the safety and economics of nuclear generators, the deregulation of natural gas, and the presence of carcinogenic agents in the workplace and in the kitchen.

A second factor compelling the use of statistics is the proliferation of data available to society. Examples are crime rates, health statistics, economic data, educational levels and achievements, and employment statistics. This growth in numbers has been accompanied by increased sophistication in the techniques used to analyze and interpret these numbers. Instead of simple tabular presentations, courts are confronted by multiple regression analyses, correlation coefficients, and tests of hypotheses.

Increased awareness on the part of attorneys poses questions as to the extent to which the law should employ statistics and the methods of inference that accompany them, what norms or guides should be established, and what aspects of statistics lawyers need to understand before they confront these pressing questions. Answers to the first two questions are beyond the scope of this work. The latter, which is the subject of this book, can perhaps be addressed initially by tracing the historical role of statistics in the litigation process, including recent innovative applications.

THE ROLE OF STATISTICS IN LITIGATION

One of the earliest reported instances of proof by statistical probabilities occurred in the infamous trial in 1899 of Alfred C. Dreyfus. To identify writing in a document as that of Dreyfus, the prosecution's witnesses reported a number of close matches between the lengths of certain words and letters in a document and the lengths of certain words and letters in correspondence taken from Dreyfus's home. As later noted by the panel of experts appointed to review the evidence in the *Dreyfus* case, there was nothing statistically remarkable about the existence of close matches in some word lengths between the disputed document and Dreyfus's correspondence. Thus, one of the first examples of introducing probabilities as proof ended on an unsatisfactory note.

The state of knowledge has advanced such that statistics are readily used in court cases today. Statistical data have been used extensively in civil rights litigation. For instance, statistics have been allowed as proof in jury discrimination,² school,³ voting,⁴ housing,⁵ and age discrimination⁶ cases. Reliance upon statistical evidence by the courts has been substantial in some instances. In *Alabama v. United States*, the court stated that "statistics often tell much and the courts listen."⁷ In *Jones v. Lee Way Motor Freight, Inc.*, the court observed that "statistics often demonstrate more than the testimony of many witnesses."⁸

Antitrust cases are another area that has relied upon statistics to establish the

relevant market as well as the defendant's share thereof. The courts generally view degree of concentration as a proxy for degree of competition in a particular market—an important element of many antitrust cases and of antitrust policy in general. In *United States v. United Shoe Machinery Corp.*, a case utilizing both descriptive and inferential statistics, the court concluded:

Through these leases United maintains a network of contracts with approximately 90 percent of all shoe factories. It supplies more than 75 percent of the demand for shoe machinery (excluding dry thread sewing machinery). From these and other facts, such as an analysis of United pricing practices, and a comparison between United and its competitors in terms of financial resources, of facilities, of accumulated experience, and of variety of machines offered, it appears that United has such substantial market power as to give it effective control of the shoe machinery market.⁹

Statistical presentations have been widely used in court cases involving deceptive advertising. In the landmark case *Bristol-Myers Co. v. FTC*, which involved inferential statistics based on sampling, the court held:

Based on the results of the survey, the corporation published advertisements from which the casual reader would reasonably infer that careful inquiry amongst the members of the dentist profession had disclosed that a large majority of the dentists in this country not only used Ipana but recommended it to their patients. . . . We are of the opinion that these sweeping statements were not justified by the answers to the questionnaire and that in consequence, as the Board found, the advertisements were misleading and likely to deceive the public.¹⁰

The most publicized use of statistical proof has occurred in Title VII cases. Courts generally have accorded statistical evidence great, if not decisive, weight. Dorsaneo asserts that the “relevance of statistical evidence is premised upon the legal principle that a statistical demonstration of the unequal consequences of a particular employment practice or selection process establishes a *prima facie* case of discrimination.”¹¹ Limitations that may have restricted the courts in their enforcement of the act appear to have been eliminated by the Supreme Court in the landmark decision *Griggs v. Duke Power Co.* The court held that “congress directed the thrust of the Act to the *consequences* of employment practice, not simply the motivation.”¹² This decision has generally resulted in the assumption that legally significant inferences can be obtained by comparing the composition of an employer's work force with that of the population from which the employees are drawn. For example, in *Hazlewood School District v. United States*, the Supreme Court held:

We noted that statistics can be an important source of proof . . . since absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population from which employees are hired . . . where gross statistical disparities can be shown, they alone may in a proper case constitute *prima facie* proof.¹³